



UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER OF PATENTS AND TRADEMARKS
Washington, D.C. 20231
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
09/421,416	10/19/1999	HINRICH SCHUETZE	D/99198	4124

7590 12/21/2001

JOHN E BECK
XEROX CORPORATION
XEROX SQUARE 20A
ROCHESTER, NY 14644

EXAMINER

FLEURANTIN, JEAN B

ART UNIT	PAPER NUMBER
2172	

DATE MAILED: 12/21/2001

Please find below and/or attached an Office communication concerning this application or proceeding.

Office Action Summary	Application No. 09/421,416	Applicant(s) Schuetze et al.
	Examiner Jean Bolte Fleurantin	Art Unit 2172

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE three MONTH(S) FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136 (a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If the period for reply specified above is less than thirty (30) days, a reply within the statutory minimum of thirty (30) days will be considered timely.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133).
- Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

- 1) Responsive to communication(s) filed on _____.
- 2a) This action is FINAL. 2b) This action is non-final.
- 3) Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11; 453 O.G. 213.
- Disposition of Claims**
- 4) Claim(s) 1-42 is/are pending in the application.
- 4a) Of the above, claim(s) _____ is/are withdrawn from consideration.
- 5) Claim(s) _____ is/are allowed.
- 6) Claim(s) 1-22, 28, and 39-42 is/are rejected.
- 7) Claim(s) 23-27 and 29-38 is/are objected to.
- 8) Claims _____ are subject to restriction and/or election requirement.

Application Papers

- 9) The specification is objected to by the Examiner.
- 10) The drawing(s) filed on _____ is/are objected to by the Examiner.
- 11) The proposed drawing correction filed on _____ is: a) approved b) disapproved.
- 12) The oath or declaration is objected to by the Examiner.

Priority under 35 U.S.C. § 119

- 13) Acknowledgement is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d).

a) All b) Some* c) None of:

1. Certified copies of the priority documents have been received.
2. Certified copies of the priority documents have been received in Application No. _____.
3. Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

*See the attached detailed Office action for a list of the certified copies not received.

- 14) Acknowledgement is made of a claim for domestic priority under 35 U.S.C. § 119(e).

Attachment(s)

- 15) Notice of References Cited (PTO-892) 18) Interview Summary (PTO-413) Paper No(s). _____
- 16) Notice of Draftsperson's Patent Drawing Review (PTO-948) 19) Notice of Informal Patent Application (PTO-152)
- 17) Information Disclosure Statement(s) (PTO-1449) Paper No(s). _____ 20) Other: _____

Art Unit: 2172

DETAILED ACTION

1. Claims 1-42 are presented for examination.

Drawings

2. The drawings filed on 10/19/1999 are approved by the Draftsperson under 37 CFR1.84 or 1.152 as indicated in the "Notice of Draftperson's Patent Drawing Review," PTO-948.

Claim Rejections - 35 U.S.C. § 102

3. The following is a quotation of the appropriate paragraphs of 35 U.S.C. 102 that form the basis for the rejections under this section made in this Office action:

A person shall be entitled to a patent unless --

(b) the invention was patented or described in a printed publication in this or a foreign country or in public use or on sale in this country, more than one year prior to the date of application for patent in the United States.

Claims 1-6, 16-17, 20-22, 28, and 39-42 are rejected under 35 U.S.C. 102(b) as being anticipated by Schuetze (US Pat. No. 5,675,819) ("Schuetze").

As per claim 1, Schuetze teaches a method for quantitatively representing objects in a vector space, as claimed comprises the steps of identifying an object to be processed from a plurality of objects (thus, a search is performed to retrieve possibly relevant documents the documents are analyzed to determine the number that are actually relevant to the query, the precision of the search is the ratio of the number of relevant documents to the number of retrieved documents; which is readable as identifying an object to be processed from a plurality of objects) (see col. 18, lines 63-67);

Art Unit: 2172

extracting a feature corresponding to the object from the plurality of objects (thus, accessing and browsing documents based on content similarity, words and documents are represented as vectors in the same multi-dimensional space that is derived from global lexical co-occurrence patterns; which is readable as extracting a feature corresponding to the object from the plurality of objects) (see col. 4, lines 9-13);

converting the feature to at least one vector (thus, in computing a document vector, those terms that correspond to the sense used in the document will be reinforced whereas the direction represented by the inappropriate sense will not be present in other words, which is readable as converting the feature to at least one vector) (see col. 8, lines 14-18);

associating the at least one vector with the object (thus, each term of the documents is associated with a vector that represents the term's pattern of local co-occurrences, this vector can then be compared with others to measure the co-occurrence similarity, and hence semantic similarity of terms; which is equivalent to associating the at least one vector with the object) (see col. 6, lines 27-32); also in column 5, lines 4 through 10, Schuetze teaches after forming the thesaurus vectors, a context vector for each document is computed, the context vector is a combination of the weighted sums of the thesaurus vectors of all the words contained in the document, these context vectors then induce a similarity measure on documents and queries that can be directly compared to standard vector-space methods.

As per claim 2, the limitations of claim 2 are rejected in the analysis of claim 1 above, and this is rejected on that basis.

Art Unit: 2172

As per claim 3, Schuetze teaches a method as claimed, wherein the feature comprises text surrounding the subject document in a host document (thus, the documents are analyzed to determine the number that are actually relevant to the query, which is readable as text surrounding the subject document in a host document) (see col. 18, lines 64-65).

As per claims 4 and 20, Schuetze teaches a method as claimed, wherein the feature comprises text represented by the subject document (thus, methods perform a computation on the text of the documents in the corpus to produce a thesaurus, which is readable as text represented by the subject document) (see col. 2, lines 17-18). Also, in column 1, lines 14 through 20, Schuetze teaches the step of the information retrieval systems typically define similarity between queries and documents in terms of a weighted sum of matching words, the usual approach is to represent documents and queries as long vectors and use similarity search techniques.

As per claims 5, 16, and 41 in addition to the discussion in claim above, Schuetze teaches the step of counting the occurrences of each unique word in the subject document (thus, the dimensionality of the thesaurus space is reduced by using a singular value decomposition the closeness of terms with equal frequency occurs because the terms have about the same number of zero entries in their term vectors, for a given term singular value decomposition assigns values to all dimensions of the space, so that frequent and infrequent terms can be close in the reduced space if they occur with similar terms, for example, the word "accident," which may occur 2590 times, and the word "mishaps," which may occur only 129 times, can have similar vectors that are close despite the frequency difference between them, the technique of singular value

Art Unit: 2172

decomposition (SVD) is used to achieve a dimensional reduction by obtaining a compact and tractable representation for search purposes, the uniform representation for words and documents provides a simple and elegant user interface for query focusing and expansion; which is readable as counting the occurrences of each unique word in the subject document) (see cols. 4-5, lines 54-3);

creating a vector having a number of dimensions equal to the number of unique words in the collection of documents, and further having as each element a numeric value representative of the number of occurrences in the subject document of the corresponding word (thus, terms are represented as high-dimensional vectors with a component for each document in the corpus, the value of each component is a function of the frequency the term has in that document they show that query expansion using the cosine similarity measure on these vectors improves retrieval performance; however, the time complexity for computing the similarity between terms is related to the size of the corpus because the term vectors are high-dimensional (see col. 3, lines 8-17).

As per claims 6, 17, and 42 Schuetze teaches a method as claimed, wherein the value representative of the number of occurrences in the subject document of the corresponding word is calculated as the token frequency weight of the corresponding word multiplied by the inverse context frequency weight of the corresponding word (thus, documents are clustered into small groups based on similarity measure two documents are considered similar if they share a significant number of terms with medium frequency terms preferentially weighted terms are then

Art Unit: 2172

grouped by their occurrence in these document clusters, since a complete-link document clustering is performed the procedure is very computationally intensive and does not scale to a large reference corpus; which is readable as wherein the value representative of the number of occurrences in the subject document of the corresponding word is calculated as the token frequency weight of the corresponding word multiplied by the inverse context frequency weight of the corresponding word (see col. 2, lines 51-57). Also, in column 17, lines 30 through 44, Schuetze teaches the step of weighting the words in the document is by using an augmented tf.idf method ‘term frequency-inverse document frequency method’ when summing thesaurus vectors: ##EQU13## where $tf.sub.ij$ is the frequency of word I in document j; N is the total number of documents; and $n.sub.i$ is the document frequency of word I. as the word frequency increases in a document, the weight (score) for that word also increases, however, the term $N/n.sub.i$ is inversely proportional to document frequency such that high frequency words receive less weight.

As per claim 21, in addition to the discussion in claim 5 above, Schuetze teaches the step of wherein the converting step comprises the steps of for each possible text genre, processing the subject document to calculate the probability that the subject document is of the corresponding genre (thus, the documents are analyzed to determine the number that are actually relevant to the query, the precision of the search is the ration of the number of relevant documents to the number of retrieved documents; which is readable as processing the subject document to calculate the probability that the subject document is of the corresponding genre) (see col. 18, lines 64-67).

Art Unit: 2172

As per claim 22, Schuetze teaches a method as claimed, wherein the feature comprises the color histogram for an image represented by the subject document (thus, truncated group average agglomerate clustering merges disjoint document sets or groups starting with individuals until only k groups remain, at each step the two groups whose merger would produce the least decrease in average similarity are merged into a single new group; which is readable as histogram for an image represented by the subject document) (see col. 10, lines 49-53).

As per claim 28, Schuetze teaches a method as claimed, wherein the feature comprises the color complexity of an image represented by the subject document (see col. 10, lines 49-53).

As per claims 39-40, Schuetze teaches a method as claimed, wherein the object to be processed comprises a subject user selected from a user population (thus, reviewing the information in tables 3 and 4, the user sees that documents 132, 14387, and 4579 are one-topic documents that are represented by words that characterize their content, documents 13609, 22872, and 27081 are long documents with more than one topic; therefore, their document vectors are closer to the global centroid their nearest neighbors are function words because function words share the characteristic of having a large number of words from different topics as their neighbors; which is readable as a subject user selected from a user population) (see cols. 13-14, lines 66-7).

Art Unit: 2172

Claim Rejections - 35 U.S.C. § 103

4. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

(a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negated by the manner in which the invention was made.

Claims 7-15 and 18-19 are rejected under 35 U.S.C. 103(a) as being unpatentable over Schuetze (US Pat. 5,675,819) in view of Li (US Pat. 5,920,859), ("Schuetze"), ("Li").

As per claims 7-10, 13-15, and 18-19 Schuetze teaches all the subject matter of the claimed invention with the exception of an exact URL representing all documents in the collection of documents; and inlinks in the collection of documents linking to the subject document; and outlinks in the subject document linking to other documents. However, Li teaches the step of the query may be represented by a query vector where the query vector contains a dimension for each term in the query, each document may be represented by document link vectors for each hyperlink pointing to the document, where each document link vector contains a dimension for each term in the corresponding hyperlink pointing to that document comparing the words in the query to the words in the hyperlinks includes calculating the dot product of the query vector with the document link vector for that hyperlink summing the relevance ranking for each hyperlink pointing to a document includes summing the dot products obtained using the document link vectors for a particular document to obtain the summed relevance score for that document, the summed relevance scores may then be compared to obtain

Art Unit: 2172

a ranking of documents; which is readable as URL representing all documents in the collection of documents; and inlinks in the collection of documents linking to the subject document; and outlinks in the subject document linking to other documents (see col. 4, lines 25-39). Thus, it would have been obvious to a person of ordinary skill in the art at the time the invention was made to modify the teachings of Schuetze and Li with the step of URL representing all documents in the collection of documents; and inlinks in the collection of documents linking to the subject document; and outlinks in the subject document linking to other documents. This modification would allow the teachings of Schuetze and Li to improve the accuracy and the reliability of the system and method for quantitatively representing data objects in vector space, and provide comparison the words in the query to the words in a hyperlink to obtain a relevance ranking for each hyperlink and summing the relevance rankings for each hyperlink pointing to a particular document to obtain a summed relevance score for that document (see col. 4, lines 19-24).

As per claim 11, the limitations of claim 11 are rejected in the analysis of claim 5 above, and this is rejected on that basis.

As per claim 12, the limitations of claim 12 are rejected in the analysis of claim 6 above, and this is rejected on that basis.

Art Unit: 2172

Allowable Subject Matter

5. Claims 23-27, 29-38 are objected to as being dependent upon a rejected base claim, but would be allowable if rewritten in independent form including all of the limitations of the base claim and any intervening claims.

6. The prior art made of record and not relied upon is considered pertinent to applicant's disclosure: Rose et al. US Pat. No. 5,870,740 relates to an information retrieval system. Corey et al. US Pat. No. 5,987,446 relates to text searching engine are utilized in searching for one or more desired information items. Deerwester US Pat. No. 5,778,362 relates to methods and systems for analyzing collections of data items to reveal structures such as associative structures within the collections of data items. Bolle et al. US Pat. No. 5,546,475 relates to the field of recognizing.

Conclusion

7. Any inquiry concerning this communication from examiner should be directed to Jean Bolte Fleurantin at (703) 308-6718. The examiner can normally be reached on Monday through Friday from 7:30 A.M. to 6:00 P.M.

If any attempt to reach the examiner by telephone is unsuccessful, the examiner's supervisor, Mrs. KIM VU can be reached at (703) 305-8449. The FAX phone numbers for the Group 2100 Customer Service Center are: *After Final* (703) 746-7238, *Official* (703) 746-7239, and *Non-Official* (703) 746-7240. NOTE: Documents transmitted by facsimile will be entered as official documents on the file wrapper unless clearly marked "**DRAFT**".

Art Unit: 2172

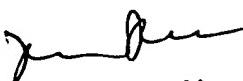
Any inquiry of a general nature or relating to the status of this application or proceeding should be directed to the Group 2100 Customer Service Center receptionist whose telephone numbers are (703) 306-5631, (703) 306-5632, (703) 306-5633.



Jean Bolte Fleurantin

December 13, 2001

JB/



HOSAIN T. ALAM
PRIMARY EXAMINER